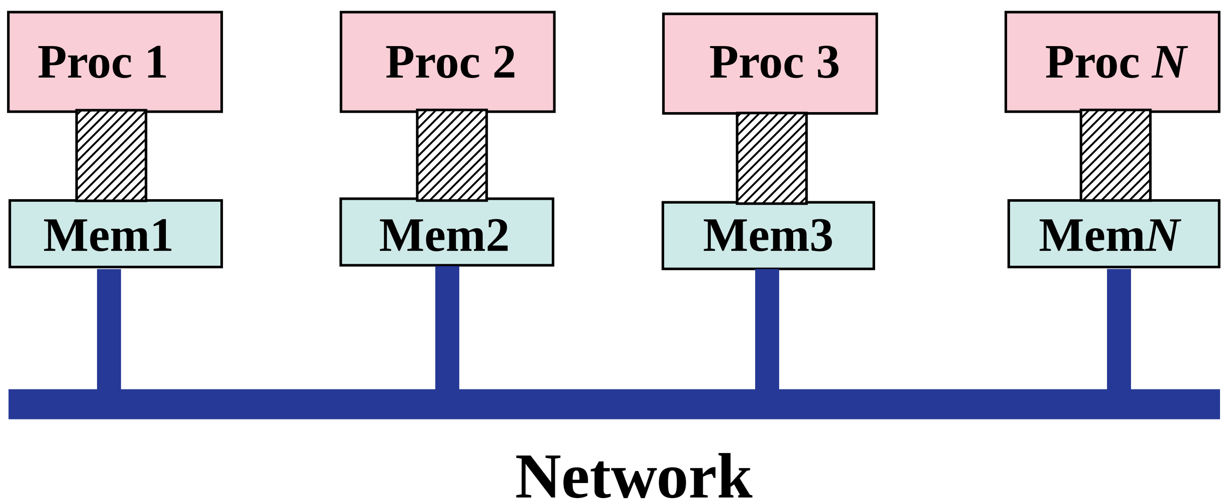
TreadMarks
Shared Memory Computing on a Network of PCs

Willy Zwaenepoel

ParallelTools LLC
and
Department of Computer Science
Rice University

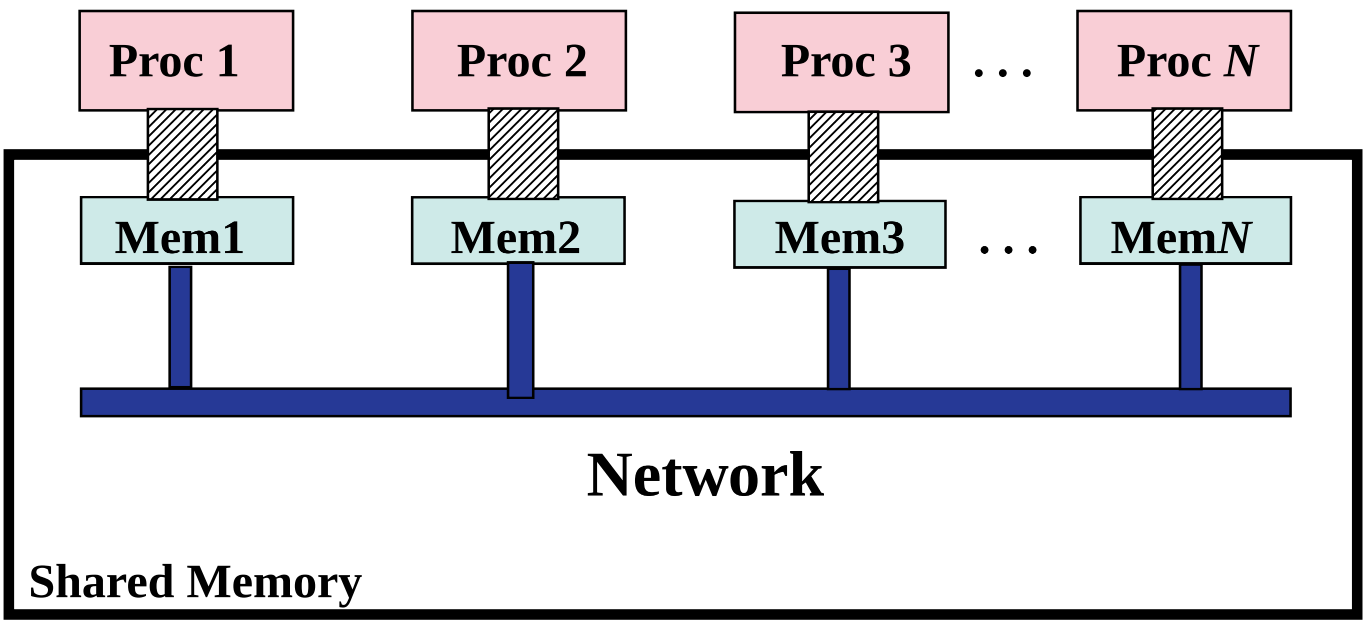
Distributed Memory Machine



Programming model: Message Passing

Distributed Shared Memory

Software provides shared memory image



Easier to program

Shared Memory API

Threads

Synchronization

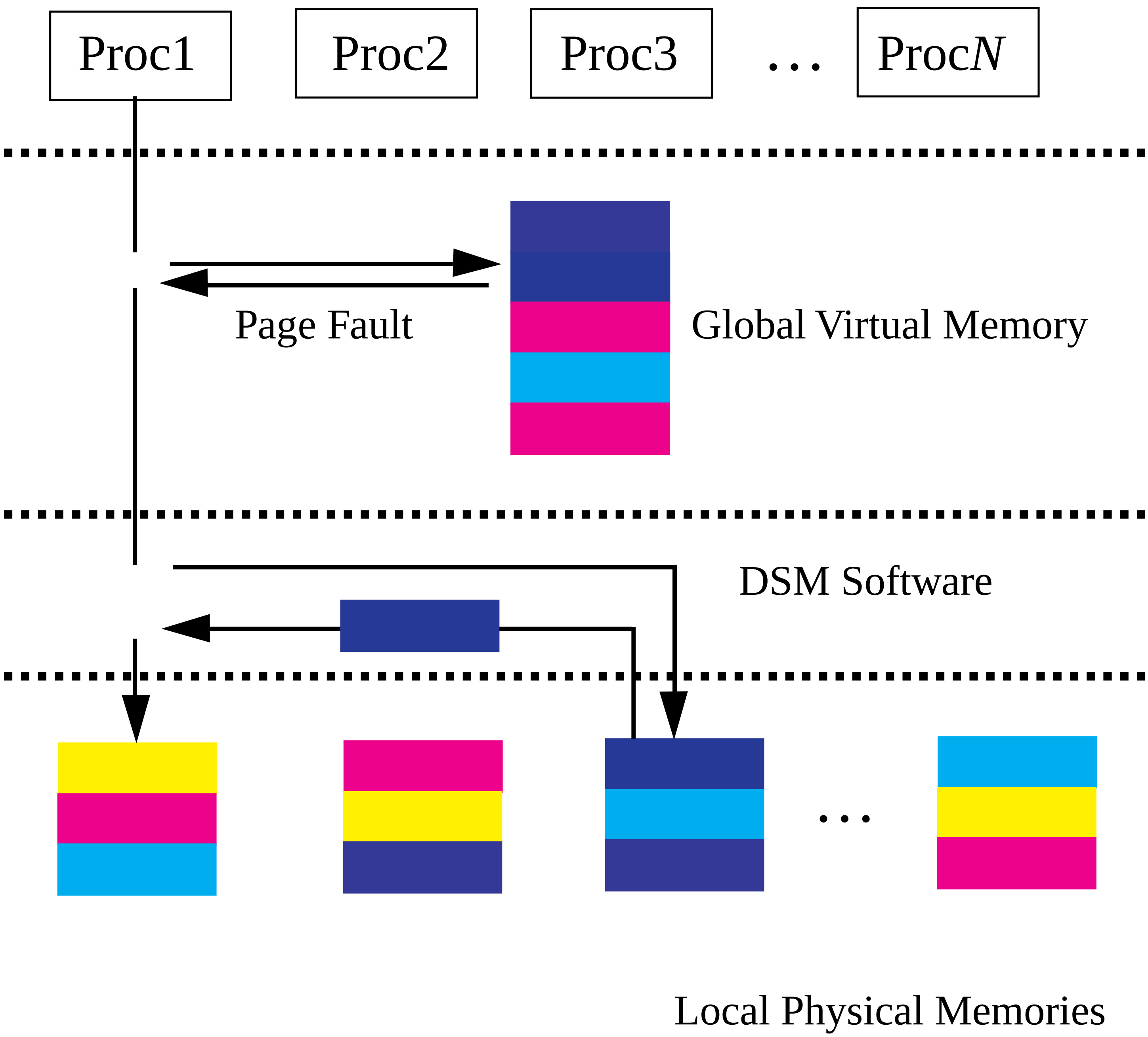
- Locks
- Barriers
- Flags

Shared memory allocation

Programming with TreadMarks

```
a = malloc ()
for( number of timesteps )
{
    for( i=0; i < n; i++ )
        for( j=0; j < n; j++ )
            t[i, j] = 0.25 * ( a[i–1, j–1] + … );
    for( i=0; i < n; i++ )
        for( j=0; j < n; j++ )
            a[i, j] = t{i, j];
}
```

Conventional DSM Implementation [Li 86]



Performance Problems and Solutions

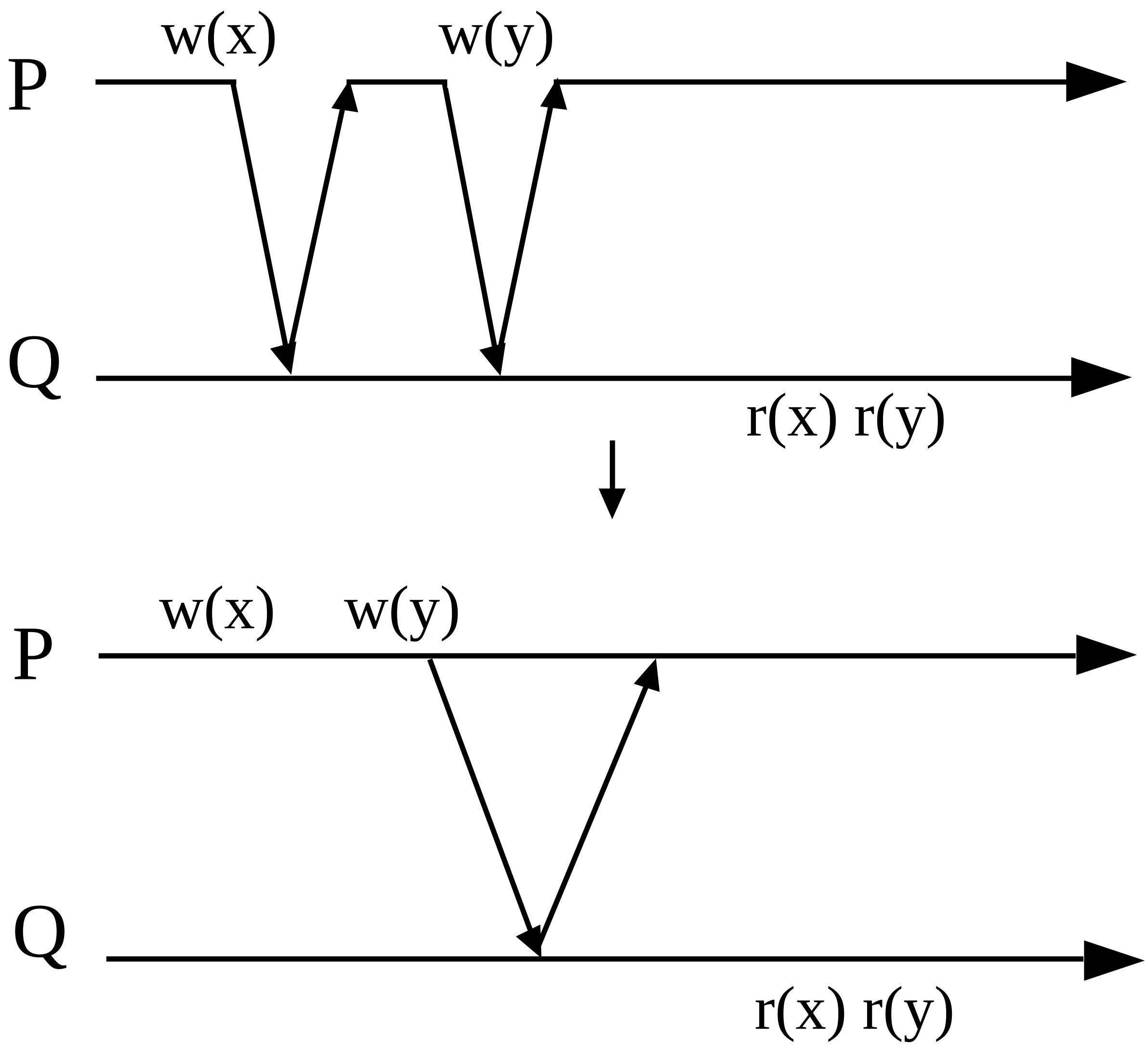
Problems: Too much communication

- Sequential consistency
- False sharing

Solutions:

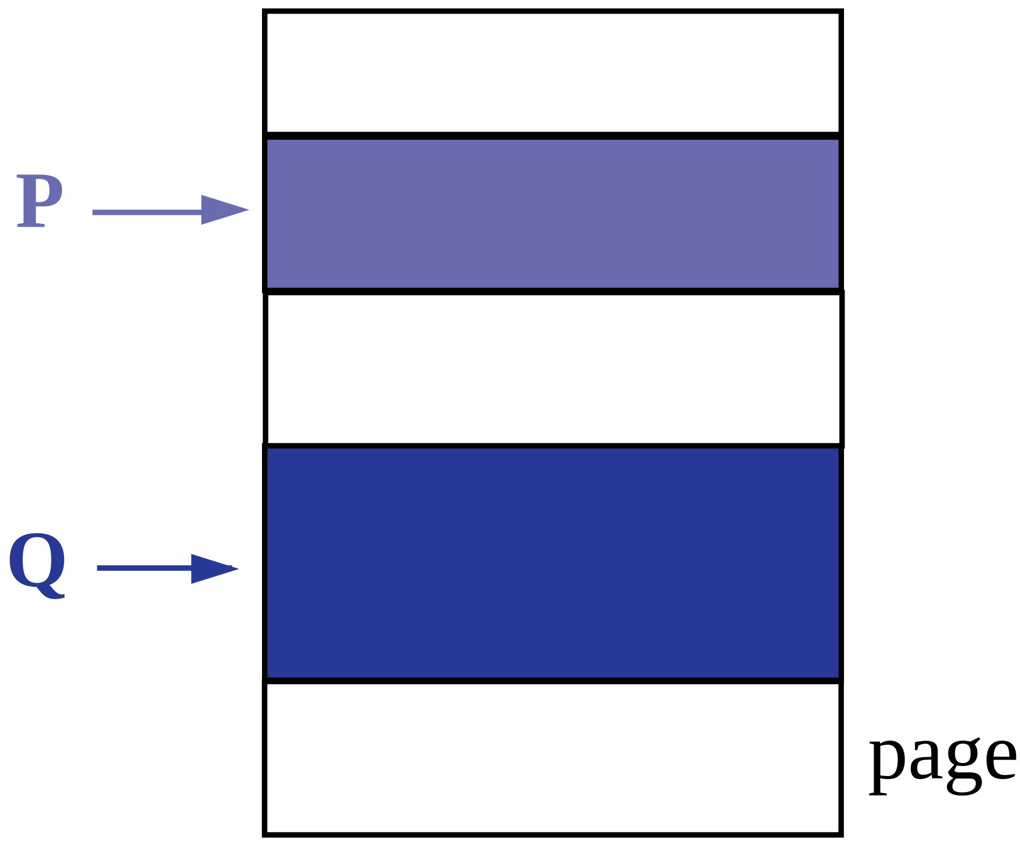
- Lazy release consistency [Keleher 92]
- Multiple writer protocol [Carter 91]

Sequential Consistency → Release Consistency

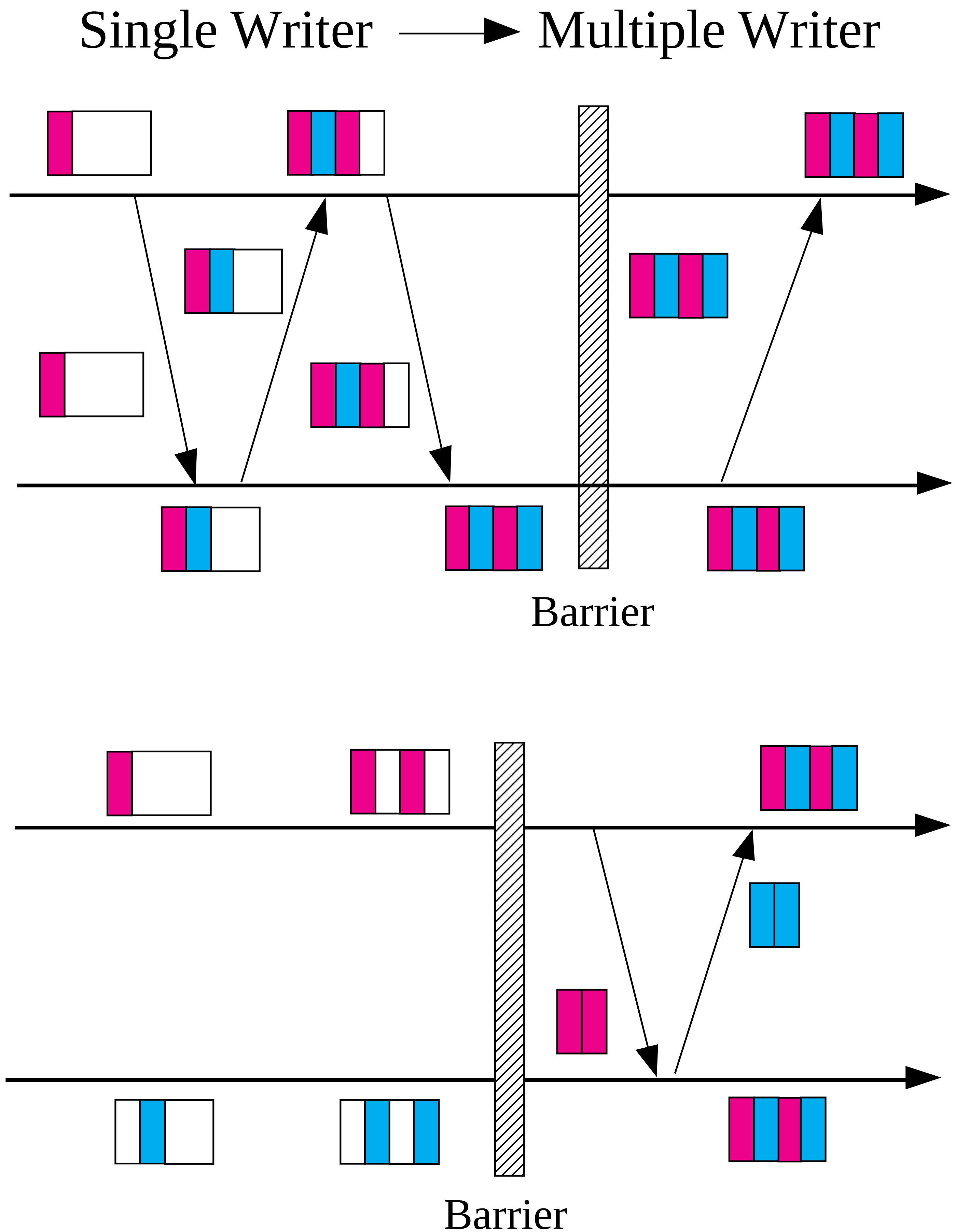


False Sharing

Pieces of the same page updated by different processors



Leads to “ping-pong” effect


Single Writer
Multiple Writer
Barrier
Barrier

Tread Marks

Standard kernel and compilers

User-level library for C and Fortran

Implemented on

- DEC
- HP
- IBM
- Intel
- Sun
- SGI

Relatively portable
[Keleher et al. 94]

Works? Doesn't work?

Highly regular codes: try using a compiler

Complicated parallelization, irregular access, pointers: try using shared memory

False sharing is often not a problem

High synchronization rate is a problem

New York Times, Nov. 15, 1996:

Scientists Identify Site of a Gene Tied to Some Parkinson’t Cases

UPI, Nov. 14, 1996:

Researchers announced they have the general location of a gene that can cause Parkinson’t disease ...

Science, Nov. 15, 1996 [Poly. et al.]:

Parkinson’t disease is the second most common neurodegenerative disorder ... Genetic markers on chromosome 4q21-23 were found ...

Genetic Linkage Analysis

Disease gene location:

- ♦ biological experiments
- ♦ computational steps (linkage analysis)

Computation is bottleneck

Hours to months is normal

Better accuracy desired

Genetic Linkage Analysis (Contd.)

Computation was doen using TreadMarks on

IBM SP2 at NIH
Network of SGI SMPs at Rice

The Linkage Computation (contd.)

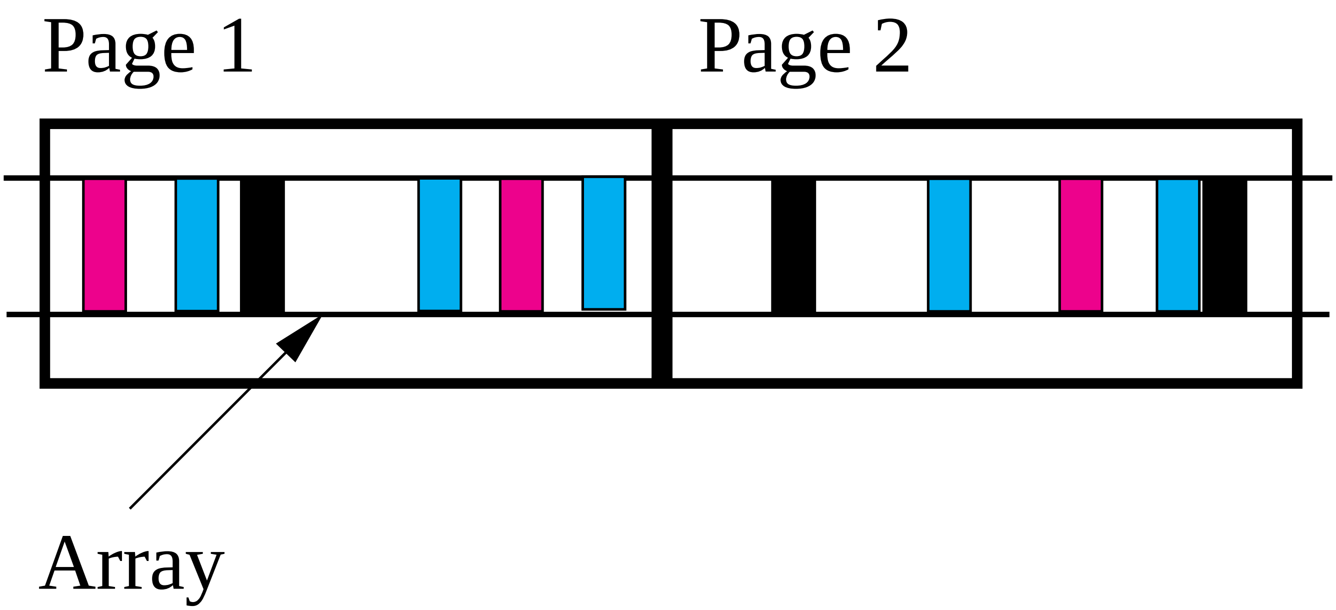
```
For a particular choice of θ
{
    Walk the tree and compute likelihood
    Compute new choice of θ
}
```

Parallelization of Linkage

For a particular choice of θ
{
    For each nuclear family
        {
            Distribute array "mostly-cyclic"
            In parallel: Compute sparse array
        }
        Compute new choice of θ
    }

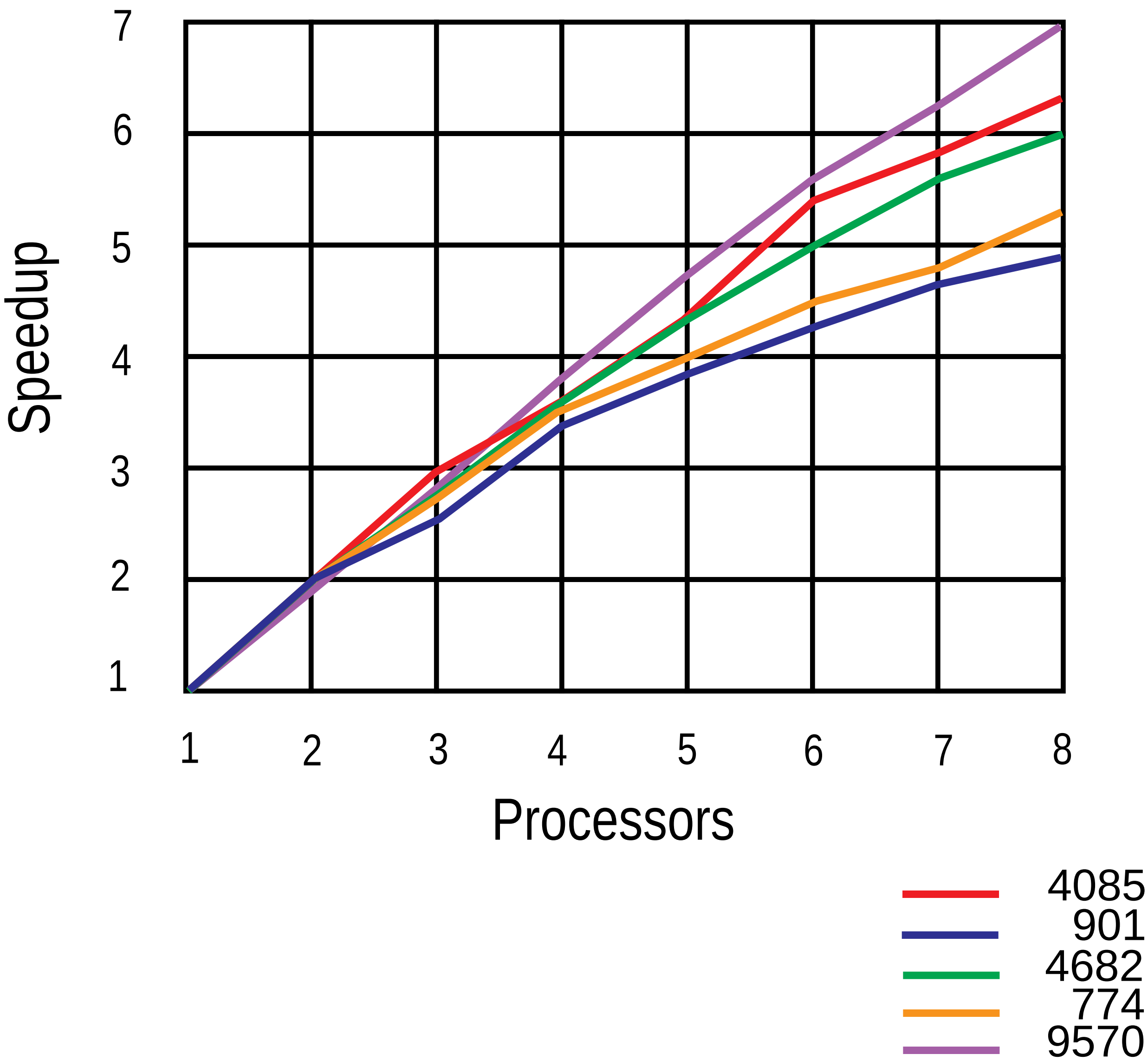
Load distribution is good, but intense false sharing

False Sharing with Cyclic Array Distribution



All processors write on every page!

Results



[Gupta et al., 1995]

Parallel FASTLINK Sites

ANGIS, Sydney, Australia (SPARC SMP)

Columbia University, New York (Alpha SMP)

Fox chase Cancer Center, Philadelphia (Alpha net work)

Griffith University, Brisbane, Australia (IBM SP-2)

Human Genome Project, Hinxton, U.K. (SGI SMP)

Infobiogen, Paris, France (SPARC SMP)

MDC für Mol. Medizin, Berlin, Germany (SPARC SMP)

NIH (IBM SP2 and Alpha network)

Ospedale San Raffaele, Milan, Italy (SPARC SMP)

Sequana Therapeutics, La Jolla (SPARC network)

University of Antwerp, Belgium (Alpha SMP)

Further Work: Tools, Tools, Tools

Compiler front-end for ease of programming

Compiler back-end for performance

Performance debugging tools

Better Programming with TreadMarks

```
for( number of timesteps )
   {
#pragma parallel pfor
      for( i=0; i < n; i++ )
         for( j=0; j < n; j++ )
            t[i, j] = 0.25 * ( a[i–1, j_1] + …);
#pragma parallel pfor
      for( i=0; i < n; i++ )
         for( j=0; j < n; j++ )
            a[i, j] = t[t, j];
   }
```

For More Information

E-mail:
treadmarks@ece.rice.edu
ptools@ns.syntron.com

WWW:
http://www.cs.rice.edu/~willy/TreadMarks/overview.html

The Linkage Computation

Maximum likelihood optimization of θ